2010-03

Team 6: Applying Social Network Analysis to Data Farming of Agent-Based Models

Meyer, Ted

Team 6: Applying Social Network Analysis to Data Farming of Agent-Based Models

TEAM 6 MEMBERS

Meyer, Ted
McDonald, Mary
Upton, Steve
Naval Postgraduate School, US

Middleton, Donna
Northrop Grumman, US

Thomas Holland
Naval Surface Warfare Center, Dahlgren, VA

Kim Lewis
Department of Defense

Harald Schaub
IABG, Germany

Fernando Freire
National Defense College, Portugal

Christina Bouwens
SAIC, US

INTRODUCTION

Team 6 continues to participate in an ongoing study to examine the utility of distillation modeling in the Counter-IED (Improvised Explosive Devices) fight. Understanding social networks, their nature in insurgencies and IED networks, and how to impact them, is important to the Counter-IED battle. Team 6 is exploring methods of extracting, analyzing, and visualizing dynamic social networks that are inherent in agent-based models in order to build tools to examine and manipulate insurgencies. We are starting with basic clique creation scenarios as the initial basis of our investigations and are examining the types of network statistics that can be used as MOEs and pointers to unique and emergent behaviors of interest.

The Team 6 goals during IDFW 20 were to extend our base scenario with simple variations and to test candidate tools and prototype methods for data farming the scenario, extracting network data, analyzing end-of-run network statistics, and visualizing network behaviors.

Social Network Analysis (SNA) techniques were explored in detail to determine which network metrics would be most beneficial for analyzing the types of networks produced by our agent based model. This would allow the team to explore questions regarding Counter-IED issues—including insurgent network evolution and adaptation. Within insurgent, IED-using networks, there are two of interest: IED Emplacement Networks (consisting of personnel that are directly involved with IED usage) and IED Enabling Networks (consisting of communities that indirectly support the IED Emplacement networks). Team 6 is in the process of identifying tools that can be used to explore patterns that might provide valuable insights into emergent behaviors of interest.

Background

In previous work, at IDFW 19 and between workshops, the team:

- Examined a set of agent-based model (ABM) C-IED (Counter-IED) task plans generated by previous workshops.
- Selected potential candidate tasks for follow–up study and analysis.
- Concluded that SNA concepts and techniques needed to be applied to address the candidate tasks.
- Demonstrated the ability to extract social network data from a basic social interaction scenario.
- Data farmed initial scenario and established need to simplify the target scenario in order to more closely examine cause and effect relationships to SNA statistics.
- Developed a new base scenario, delineated a simple illustrative DOE, and data farmed the model to provide a sample data set for further exploration.

IDFW 20 Objectives

Team 6's objectives for IDFW 20 were to:

- Examine utility and approach of applying specific SNA statistics, methods, and concepts using the data farming output provided from previous work.
- Delineate the data requirements for the various types of networks that might be extracted from modeling.
- Establish and document software and processes for applying these capabilities to detecting and analyzing emergence.

To address these objectives, the team started with a very basic approach. Assuming that an agent based simulation produces a time-series of state data and MOEs, our tools and methods need to allow the analyst to conduct tests to:

- Detect the presence of a network or networks.
- Distinguish different networks and different classes of networks.
- Determine if and when networks achieve equilibrium.
- Determine which model inputs have significant impact on the state and behaviors of the network.

Specifically, the intent is to use these capabilities to be able to address a variety of social network questions such as:

- What do insurgent networks look like? Who is in the network? Who is not?
- How do we distinguish networks that should be attacked, networks that should be attritted or that should be co-opted?
- Who are the High Value Individuals (HVI) and what are their identifiable characteristics?
- Will removing specific nodes destabilize a network?
- What are the 2nd and 3rd order effects?
- What are the potential unintended consequences?

Abstracted Illustrative Scenario and DOE

Initial work was based on the Pythagoras distribution "Peace" scenario. Data Farming of this scenario and initial analysis of the results between IDFW 19 and 20 led to the development of a more basic scenario in order to test basic network concepts.

The illustrative "Clique Creator" (CC) scenario was developed using Pythagoras's "relative" color change capability as a tool for experimenting with SNA extraction and analysis. CC has a single agent class with 100 instantiated agents that are uniformly distributed across Pythagoras's red and blue color spaces. The agents' only "weapon" is "Chat" which induces a relative color change on other agents with which the agent interacts. As the scenario is executed, entities move through various color states, becoming "more" red or "more" blue depending on the interactions with other red or blue entities. States will change depending on whether two entities engage in "chatting" and form a connection. The more any two agents interact, the more "alike" they become.

The focus of the scenario selection was to represent dynamic homophily and use the results to explore the various analysis tools under study. Multiple excursions / replications of the Pythagoras-developed Clique Creator scenario were used to produce the data for analysis with the candidate tools. This baseline provided a means for the team to experiment with various SNA measures and analysis techniques.

Pythagoras can provide multiple views of agent state data. A spatial view showed the physical relationship between entitities and where connections or bonds were formed. The inclination space view sorted the entities by colors. This color space view is used to illustrate the homophilic state of the participating entities in the simulation.

A very basic full-factorial design space was used to data farm the scenario. The design matrix (Table 1) reflects four input parameters that will influence the composition of the resulting networks:

- RelativeChange - Percentage relative change of color when "chatted."
- InfluenceRng - Maximum distance of chat.
- FriendThresh - Agents within this range are considered "linked."
- EnemyThresh – Dependent variable; is calculated as FriendThresh plus 55, in order to preserve the same Friend to Enemy Distance (equivalent to the "neutral" range) as was present in the base scenario.

Excursion	RelativeChange	InfluenceRng	FriendThresh	EnemyThresh
0	5	25	5	60
1	5	25	50	105
2	5	25	100	155
3	5	100	5	60
4	5	100	50	105
5	5	100	100	155
6	5	250	5	60
7	5	250	50	105
8	5	250	100	155
9	20	25	5	60
10	20	25	50	105
11	20	25	100	155
12	20	100	5	60
13	20	100	50	105
14	20	100	100	155
15	20	250	5	60
16	20	250	50	105
17	20	250	100	155
18	60	25	5	60
19	60	25	50	105
20	60	25	100	155
21	60	100	5	60
22	60	100	50	105
23	60	100	100	155
24	60	250	5	60
25	60	250	50	105
26	60	250	100	155

Table 1 – Clique Creator Experimental Design Matrix

The CC scenario can be considered as a metaphor for a group of people establishing relationships based on shared interests or desires (color space proximity) and physical proximity (relative agent location). Agents are drawn toward agents with similar color and move away from agents of disimilar color. The closer agents are in location, the more frequently they "chat" each other, and thus, the closer they grow in color space. Eventually, cliques of "like-interest" agent form and are impacted by other agents and cliques. The input parameters varied in the design matrix affect these behavioral processes in straightforward ways.

Visualizing the Dynamic Network State

Part of a toolset to examine social network dynamics is the ability to analyze the ongoing agent interactions, behaviors, and network responses. Co-visualizing the various aspects (layers) of network dynamics can potentially provide powerful insight into the network. Team 6 has done initial examination of the CC scenario using several visualization capabilites. Figure 1 is the spatial view provided by Pythagoras.

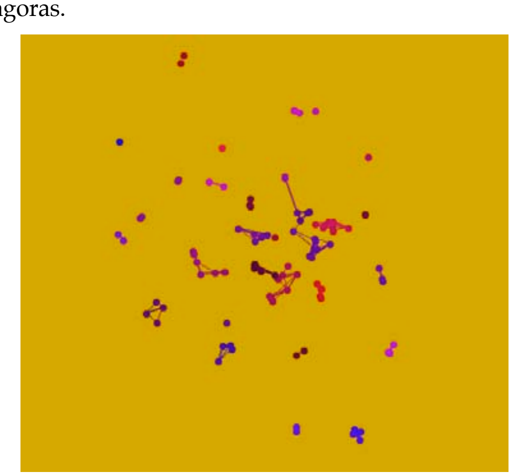

Figure 1 – Clique Creator Scenario – Spatial View

Figure 1 shows the agents at a time-step midway in the scneario. "Chats" are shown as lines between agents. This view, though, focuses on the location of the agent spatially.

Figure 2 shows four time-steps of an "inclination"-space view. In this image the location of the agents is based on their location in color space. The "redness" (0-255) of the agent is represented on the x axis. The "blueness" (0-255) of the agent is represented on the y axis. As the scenario proceeds left to right, top to bottom, note the congregation of agents into color groups. These groups do not represent the cliques formed though, because the spatial aspect is not represented.

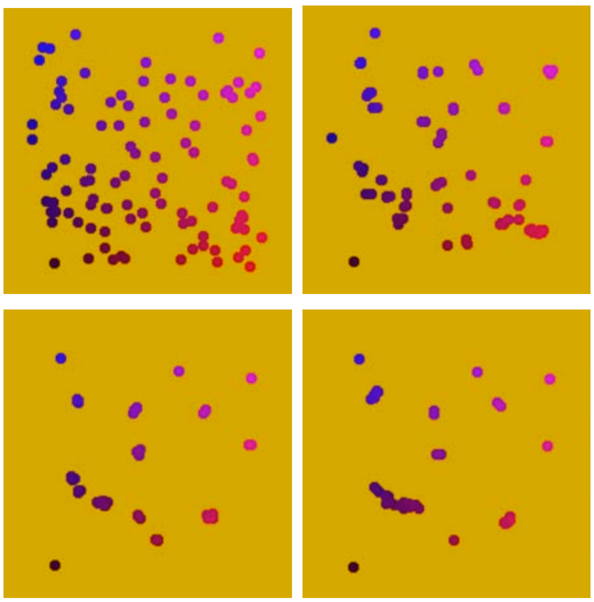

Figure 2 – Clique Creator Scenario – Inclination Space View

FIgures 3 and 4 represent the same agent network , derived from the CC scenario, using the social network analysis "layout" generated by the R SNA plug-in (http://cran.r-project.org/web/packages/sna/index.html) and SoNIA (http://www.stanford.edu/group/sonia/index.html) software packages.

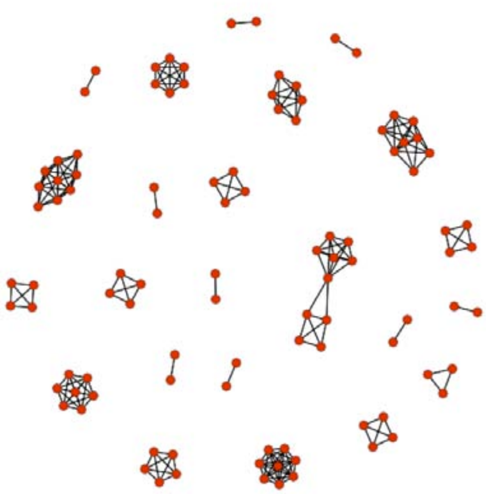

Figure 3 – Clique Creator Scenario – Static Graph View

Figure 3 shows a static network layout representation of one of the CC time-steps using the default SNA layout algorithm. The SNA R package plots each time-step independently, not accounting for the layout defined in the previous time-step. As a result, the dynamic evolution is difficult to examine.

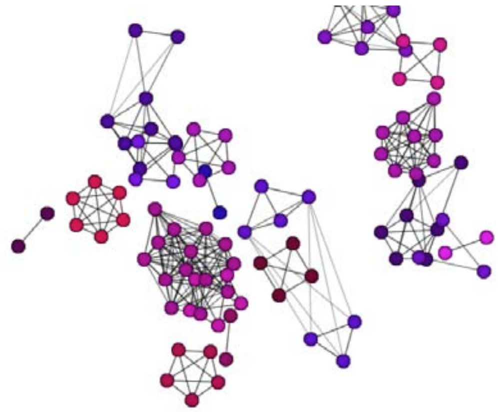

Figure 4 – Clique Creator Scenario – Dynamic Graph View

Figure 4 shows a single time-step using the SoNIA application. SoNIA is designed to support dynamic time-series network data. As a result, the layout of any timestep is based on the previous time step as a starting point. The result is a layout which displays the evolution of the network, but that can result in layouts that are not easily viewed statically.

It should be noted that Figures 2, 3 and 4 do not represent the spatial data shown in Figure 1 in any way... the "physical" location is ignored in these representations. In Figure 2 location represents color, and in Figures 3 and 4 the location is purely a function of the layout algorithm, which is designed to display the network in an uncluttered and easily-viewed manner, not the spatial location of the agents.

Social Network Analysis (SNA)

One of Team 6's goals is to begin to understand the utility of various SNA statistics in understanding the scenario dynamics and the result of data farming. Step one in this process during this workshop was to delineate what outputs and analysis methods provide insight into network evolution and impact on agent behaviors.

SNA statistics fall into two classes: node statistics and network statistics. Node statistics include: betweenness, closeness, eigenvector centrality, and degree. Network statistics include: number of components, number of cliques, and average path length.

The team decided to focus on node statistics initially and produced time-series output for every node of betweenness, eigenvector centrality and degree. Although data for 27 excursions of data farming was collected, it was decided to do an intial comparison of three excursions, where the primary variation was the color distance that defined what is considered a friend (a homophilic link) . Excursions 1, 2, and 3 were examined.

Figures 5a, 5b, and 5c represent a single replication of excusions 0, 1, and 2 as delineated in Table 1. The plots represent the degree of each agent over time. The vertical axis is degree (the number of links associated with a node), the horizontal axis is time, and the axis going into the page is agent number. Figure 5 was generated using the PlotGL plugin to R.

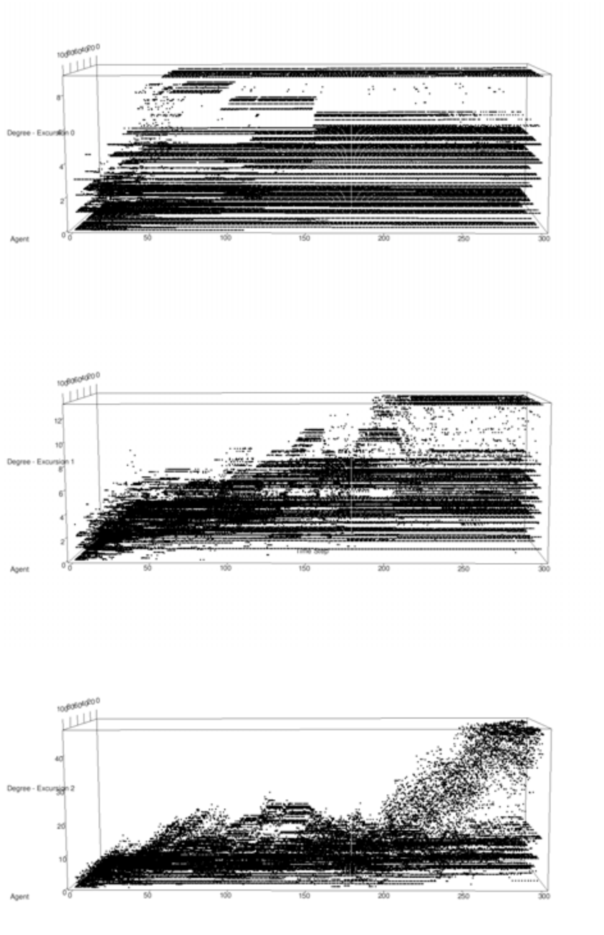

Figure 5 – Degree Centrality for Excursions 1-3

In Figure 5, various pattern differences, related to the evolution and devolution of cliques and components, can be discerned There are obvious differences between the excusions, with 0 and 1 appearing to reach covergence, but 2 never converging. It can be seen that some agents reach a steady-state and maintain it for some time, while other groups of agents particpate in behaviours which lead to the growth and reduction of degree for groups of agents.

Surprises

Two surprises (counter-intuitive results) presented themselves. Excursion 2, in Figure 5c, shows that an increase in FriendThresh, that is, expanding the range and number of agents that an agent has homophilic links with in color space leads to increased instability in terms of clique formation. The initial assumption was that this would affect the size of the cliques and number of components. The unexpected result is that this increase prevents the stabilization of cliques and network components. Rather, it appears that this increase results in groups being able to “steal” members from other groups more easily.

Another interesting behavior is the Excursion 0 (Figure 5a) degree variation that occurs before equilibrium. In this case it appears that larger components are formed intially, but that they devolve into smaller groups over time. The team intends to investigate the set of replicates associated with this excursion to determine whether this behavior is consistent for this level of FriendThresh.

Summary and Way Ahead

Significant insight was gained by team members in delineating capabilities needed in a toolkit for the extraction and analysis of dynamic social data from models. The following capabilities will be needed for ongoing data farming research of basic social networks:

- Synching of Visualization: Various representations of the dynamic network are useful, but examining multiple views of the network time-step synced would provide powerful relational insights.
- Equilibrium Time: Determining whether equilibrium occurs and how long it takes is often the first step in analysis.
- Data Farm Time Window Reduction Size: Dynamic network analysis requires defining what constitutes a link, for example, a single interaction or multiple interactions over some time window. Being able to data farm this time window would provide analysts insight into network basics.
- Node Statistic Capability: Degree, betweenness, eigenvector, closeness need to be extractable for each node, time-step, replicate and excursion and then represented effectively.
- Network/Component Statistic Capability: # cliques, and components, density, and others need to be acquired for each time step, replicate and excursion.
- Newcomer/Leaving Effects: Measure the effects of dynamic birth and death of agents.
- Network Boundary Effects: Data farm the impact of varying the size and extent of the network.
- MOEs (end-of-run vs. time-series).

Team 6 will continue to delineate tool capabilities for data farming social network models. We intend to accomplish the folowing tasks in the upcoming months:

- Document tools and methods identified in IDFW20.
- Define model output requirements for SNA analysis.
- Expand toolkit to include additional network, node, and link statistics.
- Expand data farming methods for other network layers including weapon and resource interaction, spatial, communication, and multiple “inclination” parameters.
- Continue detailed analysis of CliqueCreator data farming results.
- Test use of tools and methods on other models (MANA, Netlogo scenarios).
- Begin delineating insurgent IED network scenario.